PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism

Garrett B. Goh,[1] Benjamin S. Hulbert,[1] Huiqing Zhou,[1] and Charles L. Brooks III[1,2]*

[1] Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109

[2] Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109

ABSTRACT

pH is a ubiquitous regulator of biological activity, including protein-folding, protein-protein interactions, and enzymatic activity. Existing constant pH molecular dynamics (CPHMD) models that were developed to address questions related to the pH-dependent properties of proteins are largely based on implicit solvent models. However, implicit solvent models are known to underestimate the desolvation energy of buried charged residues, increasing the error associated with predictions that involve internal ionizable residue that are important in processes like hydrogen transport and electron transfer. Furthermore, discrete water and ions cannot be modeled in implicit solvent, which are important in systems like membrane proteins and ion channels. We report on an explicit solvent constant pH molecular dynamics framework based on multi-site λ-dynamics (CPHMD$^{MS\lambda D}$). In the CPHMD$^{MS\lambda D}$ framework, we performed seamless alchemical transitions between protonation and tautomeric states using multi-site λ-dynamics, and designed novel biasing potentials to ensure that the physical end-states are predominantly sampled. We show that explicit solvent CPHMD$^{MS\lambda D}$ simulations model realistic pH-dependent properties of proteins such as the Hen-Egg White Lysozyme (HEWL), binding domain of 2-oxoglutarate dehydrogenase (BBL) and N-terminal domain of ribosomal protein L9 (NTL9), and the pK_a predictions are in excellent agreement with experimental values, with a RMSE ranging from 0.72 to 0.84 pK_a units. With the recent development of the explicit solvent CPHMD$^{MS\lambda D}$ framework for nucleic acids, accurate modeling of pH-dependent properties of both major class of biomolecules—proteins and nucleic acids is now possible.

Proteins 2014; 82:1319–1331.

Key words: CPHMD; pK_a values; protein electrostatics; λ-dynamics; pH; molecular dynamics; protein dynamics; explicit solvent.

INTRODUCTION

pH is one of the critical regulators of biological activity. Enzymatic activity is optimized within a narrow pH range,[1] often requiring the participation or presence of ionizable residues such as aspartic acid, glutamic acid and/or histidine in the active site,[2] and accurate measurement of their pK_a values is crucial in understanding the catalytic mechanism.[3–5] In recent years, the role of pH regulation in nucleic acid systems has been acknowledged,[6,7] where parallels to proteins can be drawn, such as the catalytic activity of ribozymes (ribonucleic acid enzyme),[8–11] demonstrating the ubiquity of pH regulation in biological processes. Apart from its influence on catalytic activity, pH regulation has been observed in numerous other processes including protein folding,[12–15] protein-protein interactions,[16] protein-substrate binding,[17,18] translational recoding,[19] and aberrant pH regulation has even been implicated in cancer-related physiology.[20] As such, specific examples of pH-dependent properties encompass a wide variety of systems, such as the catalytic mechanism of dihydrofolate reductase,[21] proton gradient driven ATP synthesis,[22] activity of the U6 intramolecular stem-loop of the spliceosome complex[23] and the influenza virus haemagglutinin.[24]

While the pK_a values of amino acid monomers have been known for decades, the microenvironment around the residue located in a protein environment may alter

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: National Institutes of Health; Grant numbers: GM037554 and GM057513; Grant sponsor: Interdisciplinary Opportunities in Biophysics REU at the University of Michigan; Grant number: PHY-1004618.
*Correspondence to: Charles L. Brooks III, Department of Chemistry, 930 N. University, Ann Arbor, MI 48109. E-mail: brookscl@umich.edu
Received 11 February 2013; Revised 13 November 2013; Accepted 12 December 2013
Published online 19 December 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24499

its pK_a value. Thus, the ability to measure the microscopic pK_a of a site-specific residue is invaluable in identifying key titrating residues and understanding the mechanism of these pH-dependent biological processes. Using staphylococcal nuclease and its mutants as a model system, Garcia-Moreno and co-workers undertook a series of comprehensive investigations into the effect that the local microenvironment has on the perturbation of the pK_a values of protein residues.[25–28] Their effort has cumulated in a joint collaboration between experimentalists and theoreticians, where the current state-of-the-art computational methods for predicting protein pK_a values were evaluated against experimentally measured pK_a values.[29,30]

One major physics-based approach that has emerged over the years to treat electrostatics in proteins and nucleic acids is the Poisson-Boltzmann (PB) equation methodology, which has achieved reasonable success in predicting protein pK_a values.[31] A key limitation of initial PB methods was the lack of conformational flexibility, although this has been partially addressed using approaches like tuning the effective protein dielectric constant[32] and including representations of multiple conformations.[33,34] The need for conformational flexibility led to the development of the other major physics-based approach in computational pK_a predictions, which is based on molecular dynamics (MD) simulation. Warshel and co-workers were the first to demonstrate the use of free energy calculations to calculate the pK_a values of protein residues.[35–38] Subsequent developments in the MD community have sought to couple the protonation state of the titrating residue with the dynamics of the protein itself. Such pH-coupled simulations, which have been termed constant pH molecular dynamics (CPHMD), are uniquely suited to model realistic pH-dependent responses, even in systems where there is limited experimental data because no *a priori* information on the identity of key titrating residues and their protonation state is required.

The CPHMD methodology has been implemented using two distinct approaches, which vary in the manner in which the titration coordinates are treated—either discretely or continuously.[39] In the discrete CPHMD variant, the MD sampling of atomic coordinates is combined with the Monte Carlo (MC) sampling of protonation states. At regular intervals during a typical MD simulation, a MC step is performed to determine the change of the protonation state. Discrete CPHMD was first reported by Burgi *et al.*,[40] which was computationally expensive at that time and suffered from convergence issues, owing to the fact that it was performed in explicit solvent and used the more expensive thermodynamic integration approach to calculate the energies used in the MC evaluation step. Baptista *et al.* reported a similar discrete CPHMD implementation but used the Poisson-Boltzmann finite-difference method to calculate the energies used in the MC evaluation step.[41–43] With the advances in implicit solvation models around this time,[44,45] and the initial convergence issues reported for explicit solvent CPHMD,[40] subsequent developments in discrete CPHMD by Dlugosz and Antosiewicz,[46,47] and Mongan *et al.*,[48] were implemented using a Generalized-Born (GB) implicit solvent model. More recent improvements in the discrete CPHMD community have been focused on achieving better sampling by enhanced sampling techniques, such as Accelerated Molecular Dynamics by Williams *et al.*[49] and replica exchange strategies by Roitberg and co-workers.[50–52] Others in the field, namely Messer *et al.* have focused on developing a more physically realistic form of CPHMD,[53] using time-dependent MC sampling of the proton transfer process,[54] and the empirical valence bond (EVB) framework to simulate proton transfer between solute and solvent.[55]

By contrast, in the continuous CPHMD variant, which were first reported by Baptista *et al.*[56] and Borjesson *et. al.*,[57] titration coordinates can be treated as mixed states. In the continuous CPHMD variant developed by Brooks and co-workers, the titration coordinate represents an instantaneous microstate, and it is propagated continuously between the protonated and unprotonated states using the λ dynamics approach.[58–60] Continuous CPHMD allows one to avoid sudden jumps in potential energy that occur after a successful MC move in the discrete CPHMD variant, and potentially avoids artifacts that may be caused by the MC moves in titration coordinates. Additionally, continuous CPHMD facilitates coupled proton moves, which would need to be engineered as specific move types in the MC-based variant. Continuous CPHMD was originally implemented in implicit solvent,[61] improved to account for proton tautomerism,[62] and it provided the first demonstration of using enhanced sampling strategies to accelerate sampling and convergence in CPHMD simulations.[63] The effectiveness of continuous CPHMD has been demonstrated on numerous pH-dependent systems, inclusive of protein folding,[64,65] aggregation of Alzheimer's beta-amyloid peptides,[66] pH-triggered chaperon activity of HdeA dimers,[67] electrostatic effects on protein stability,[68] self-assembly of spider silk proteins,[69] and RNA silencing in the carnation italian ringspot virus.[70] Other investigators in the field have also seen a number of successes using discrete CPHMD simulations.[71–73]

While the move to implicit solvent CPHMD has obvious advantages in sampling and convergence, a number of unresolved issues have emerged over the years. It has been reported that the generalized Born implicit solvent model underestimates the desolvation of buried charge-charge interactions,[63] causing a systematic overstabilization of the ionized form[74] and consequently increasing the error of predicted pK_a values. In addition, these models are known to cause structural compaction which may distort the overall structure,[68,75] introducing another source of error to the pK_a calculations. Furthermore, in systems such as ion channels[76–78] and some transmembrane proteins,[79] where the microscopic interactions of discrete ions and water with the protein are important, the use of an explicit solvent representation of the solvent environment is desirable. Thus, recent developments in the continuous CPHMD community have been focused on re-introducing explicit solvent into the CPHMD framework. Wallace and Shen were the first to report a hybrid solvent CPHMD model, and showed that using an explicit solvent representation of the protein's conformational dynamics can reduce the errors introduced by the GB implicit model.[75] Around the same time, the first "pure" explicit solvent CPHMD was reported by Donnini *et al.*, with a proof of concept demonstration for model amino acid compounds.[80] Brooks and co-workers subsequently reported another "pure" explicit solvent constant pH MD simulation, termed as CPHMD$^{MS\lambda D}$ as it is based on the newer multi-site λ-dynamics (MSλD) framework.[81,82] CPHMD$^{MS\lambda D}$ was developed initially for investigating pH-dependent behavior of nucleic acid systems and it has been validated on both model nucleotides[83] and larger RNA systems.[84] The initial challenges associated with convergence in explicit solvent were noted by practitioners in the field,[80,83] and have been addressed to some extent using enhanced sampling strategies.[75,85] More recent developments in enhanced sampling methods, such as Orthogonal Space Random Walk by Zheng *et al.* have demonstrated that accurate pK_a calculations for buried protein residues in explicit solvent simulations can be achieved,[86] indicating that practical challenges first encountered by Burgi *et al.* over a decade ago will be resolved eventually.

In this article, we will adopt the explicit solvent CPHMD$^{MS\lambda D}$ framework and extend its application to include proteins. We apply the multi-site λ-dynamics (MSλD) algorithm to seamlessly perform alchemical reactions between protonation and tautomeric states, and develop a novel biasing potential to ensure that the physical end-states are adequately sampled. The quality of the CPHMD$^{MS\lambda D}$ model for proteins will be demonstrated by its ability to reproduce the pK_a values of model compounds, simulate coupled pH-dependent behavior of dipeptides, and to accurately reproduce experimental pK_a values of proteins, such as the hen egg-white lysozyme (HEWL), the binding domain of 2-oxoglutarate dehydrogenase (BBL) and the N-terminal of ribosomal protein L9 (NTL9).

THEORY

Constant pH molecular dynamics simulations in explicit solvent

We briefly review the theory behind constant pH molecular dynamics (CPHMD). In CPHMD, the protonation state of the titrating residue is described by a continuous variable, λ. The dynamics of λ is described according to multi-site λ-dynamics (MSλD),[81,82] a formalism that couples the dynamics of λ to the dynamics of the protein system. The simulation is under the influence of a hybrid Hamiltonian and its potential energy is described by:

$$U_{\mathrm{tot}}(X,\{x\},\{\lambda\})=U_{\mathrm{env}}(X)+\sum_{\alpha=1}^{N_{\mathrm{sites}}}[\lambda_{\alpha,1}(U(X,x_{\alpha,1})) \\ +\lambda_{\alpha,2}(U(X,x_{\alpha,2}))] \qquad (1)$$

where N_{sites} is the total number of titrating residues, X represents the coordinates of the environment atoms (i.e., the parts of the protein that are not titrating), and $x_{\alpha,1}$ and $x_{\alpha,2}$ represent the coordinates of atoms in residue α that are associated with the protonated and unprotonated states, respectively. λ serves as a scaling factor that is associated with each titrating residue α and its value describes the physically relevant protonated ($\lambda_{\alpha,1}=1$) and unprotonated ($\lambda_{\alpha,2}=1$) states. Details about the theoretical and methodological treatment of λ dynamics can be found in the following references.[81,82]

CPHMD simulations are calibrated on model compounds (i.e., amino acids) to reproduce the external pH environment. Modeling of the external pH is achieved by introducing a fixed biasing potential parameter ($\mathrm{F}_{\alpha,2}^{\mathrm{fixed}}$) to the unprotonated state, which results in the biased potential energy function:

$$U_{\mathrm{tot}}(X,\{x\},\{\lambda\})=U_{\mathrm{env}}(X)+\sum_{\alpha=1}^{N_{\mathrm{sites}}}[\lambda_{\alpha,1}(U(X,x_{\alpha,1})) \\ +\lambda_{\alpha,2}(U(X,x_{\alpha,2})-F_{\alpha,2}^{\mathrm{fixed}})] \qquad (2)$$

The free energy of protonation ($\Delta\mathrm{G}_{\mathrm{protonation}}$) is used to calibrate the biasing potential ($F_{\alpha,2}^{\mathrm{fixed}}$) that simulates the effect of an external pH environment. By setting the value of $F_{\alpha,2}^{\mathrm{fixed}}$ to $\Delta\mathrm{G}_{\mathrm{protonation}}$, approximately equal populations of protonated and unprotonated states are sampled in the simulation. Under this condition, the external pH environment is equal to the pK_a value of the model compound. To change the pH of the simulation, $F_{\alpha,2}^{\mathrm{fixed}}$ can be adjusted by the following equation:

$$F_{\alpha,2}^{\mathrm{fixed}}=\Delta\mathrm{G}_{\mathrm{protonation}}+\ln(10)k_BT(\mathrm{pK}_a-\mathrm{pH}), \qquad (3)$$

where pH is the external pH of the simulation and pK_a is the experimental pK_a of the model compound. The fixed biasing potential is pre-calculated and its value, corresponding to the specified external pH, is universally applied to all residues of the same type regardless of the protein environment it is in. In explicit solvent CPHMD simulations, when the titration coordinates are allowed to propagate dynamically, the two end points that correspond to physical protonation states may not be sufficiently sampled to yield converged estimates of the pK_a shifts. To ameliorate this issue, the inclusion of an extra variable biasing potential (F^{var}) is introduced, which can be adjusted to tune the sampling efficiency of titration coordinates and the fraction of physical protonation states:

$$F^{\text{var}}_{\alpha,i} = \begin{cases} k_{\text{bias}}\left(\lambda_{\alpha,i}-0.8\right)^2; & \text{if } \lambda_i < 0.8 \\ 0; & \text{otherwise} \end{cases} \tag{4}$$

Thus, in the CPHMD treatment, titratable groups in proteins may be viewed as model compounds that are perturbed by the introduction of the protein environment. Further details on the implementation and calibration of explicit solvent CPHMD$^{\text{MS}\lambda\text{D}}$ can be obtained from the following reference.[83]

pH replica exchange sampling protocol

The potential for slow convergence of protonation state sampling in CPHMD simulations has been well documented, and is exacerbated for residues with conformationally-coupled pK_a values, where they undergo a local conformation change that causes them to sample different electrostatic environments yielding distinct microscopic pK_a values.[49,84,87] Early work by Khandogin and Brooks on protein CPHMD simulations has demonstrated that the introduction of a temperature replica exchange (T-REX) protocol can significantly accelerate sampling to address such issues.[63] However, using T-REX in explicit solvent MD simulations typically incurs a large computational expense, for example, a moderate sized protein of ∼100 residues (40k atoms when solvated) requires at least 20 replicas to achieve reasonable exchange rates between adjacent temperature replicas, and when simulating CPHMD across a reasonable pH range (e.g., pH 5 to 9), the total number of replicas required increases to ∼100. Therefore, in this paper, we used a pH replica exchange (pH-REX) sampling strategy instead, and the pH-REX sampling protocol implemented in our work is based on the work of Wallace and Shen,[75] where simulations performed at various pH conditions are exchanged based on the following Metropolis criterion:

$$P = \begin{cases} 1; & \text{if } \Delta \le 0 \\ \exp(-\Delta); & \text{otherwise} \end{cases} \quad \text{where} \quad \Delta = \beta \begin{bmatrix} U^{\text{pH}}(\{\lambda_i\};\text{pH}') + U^{\text{pH}}(\{\lambda_i{}'\};\text{pH}) \\ -U^{\text{pH}}(\{\lambda_i\};\text{pH}) - U^{\text{pH}}(\{\lambda_i{}'\};\text{pH}') \end{bmatrix} \tag{5}$$

where β is $1/k_bT$, the first two terms, $U^{\text{pH}}(\{\lambda_i\};\text{pH}')$ and $U^{\text{pH}}(\{\lambda_i{}'\};\text{pH})$ are the pH-biasing potential energies for the two adjacent replicas after the exchange, and the next two terms, $U^{\text{pH}}(\{\lambda_i\};\text{pH})$ and $U^{\text{pH}}(\{\lambda_i{}'\};$ pH')are the corresponding energies for the respective replicas before the exchange.

METHODS

Generating input structures

Input structures of the peptide compounds were generated from the CHARMM topology files using the *IC* facility in CHARMM.[88] The input structure for the protein hen egg-white lysozyme (HEWL), the 45-residue binding domain of 2-oxoglutarate dehydrogenase multienzyme complex (BBL) and the 56-residue N-terminal domain of ribosomal protein L9 (NTL9) were generated from the PDB file (accession codes: 2LZT, 1W4H, 1CQU, respectively). Hydrogen atoms were added using the *HBUILD* facility in CHARMM. Model compounds (single amino acids), test compounds (dipeptide sequences) and the proteins were solvated in a cubic box of explicit TIP3P water[89] using the convpdb.pl tool from the MMTSB toolset.[90] For each protein, the system was first neutralized, before an appropriate number of Na^+ and Cl^- counterions was added to match the experimental ionic strength of 100 m*M* NaCl. All systems were capped at the N-terminus and C-terminus using CHARMM's *ACE* and *CT2* patches. Additional patches were constructed to represent the protonated forms of Asp, Glu, His, and Lys. All of the associated bonds, angles and dihedrals were explicitly defined in the patch. Each titratable residue was simulated as a hybrid model that explicitly included atomic components of both the protonated and unprotonated forms. The CHARMM parameters for the partial charges of aspartic acid, glutamic acid and lysine used in this study were reported previously by Lee *et al.*[61] Partial charges for the three protonation states of histidine were obtained without modification from the HSP, HSE and HSD residues as reported in the CHARM22 all-atom force field for proteins.[91]

Simulation details

MD simulations were performed within the CHARMM macromolecular modeling program (version c36a6) using the CHARMM22 all-atom force field for proteins[91] and TIP3P water.[89] The SHAKE algorithm[92] was used to constrain the hydrogen-heavy atom bond lengths. The

Leapfrog Verlet integrator was used with an integration time step of 2 fs. A non-bonded cutoff of 12 Å was used with an electrostatic force shifting function and a van der Waals switching function. While group-based 8 Å cutoffs investigated in the 1990s were notoriously poor in reproducing accurate dynamics of biomolecules relative to the Ewald summation technique,[93,94] modern atom-based cutoff schemes with sufficiently long cutoff distances (12 Å),[95] such as those employed in this study, has been shown to be comparable to the Ewald summation technique in modeling the dynamics of both proteins[96] and nucleic acids.[97] Titration was performed in the multi-site λ-dynamics framework (MSλD)[81,82] within the BLOCK facility, using the λ^{Nexp} functional form for λ (*FNEX*) with a coefficient of 5.5.[81,82] The titratable fragment included the protonation site and adjacent atoms whose partial charge differed according to the protonation state. The environment atoms were defined as all atoms that were not included in the titratable fragments. Linear scaling by λ was applied to all energy terms except bond, angle and dihedral terms, which were treated at full strength regardless of λ value to retain physically reasonable geometries. Each θ_α was assigned a fictitious mass of 12 amu•Å^2 and λ values were saved every 10 steps. The temperature was maintained at 298K by coupling to a Langevin heatbath using a frictional coefficient of 10 ps^{-1}. After an initial minimization, the system was heated for 100 ps and equilibrated for 100 ps (for amino acid compounds and dipeptides) or 400 ps (for HEWL, BBL, and NTL9). This was followed by a production run of 3 ns (for amino acid compounds, dipeptides and NTL9), 5 ns (for BBL), or 20 ns (for HEWL). CPHMD$^{\text{MS}\lambda\text{D}}$ simulations were performed across the pH range, with integer value pH spacing, as indicated in the titration curves (see Results and Discussion), from pH 1 to 7 for Asp and Glu, pH 4 to 9 for His, pH 7 to 12 for Lys, pH 0 to 8 for HEWL, and pH 0 to 5 for BBL and NTL9. In the pH-REX simulations, exchange attempts were made at every 1 ps. All CPHMD$^{\text{MS}\lambda\text{D}}$ simulations were performed in triplicate.

Calculation of pK_a value

The populations of unprotonated (N^{unprot}) and protonated (N^{prot}) states are defined as the total number of times in the trajectory where conditions $\lambda_{\alpha,1} > 0.8$ and $\lambda_{\alpha,2} > 0.8$ are satisfied respectively. They are used in the calculation of the fraction of physical states, which is the ratio of N^{unprot} and N^{prot} states over all states (which include intermediate λ values). The unprotonated fraction (S$^{\text{unprot}}$) is calculated for each pH window:

$$S^{\text{unprot}}(\text{pH}) = \frac{N^{\text{unprot}}(\text{pH})}{N^{\text{unprot}}(\text{pH}) + N^{\text{prot}}(\text{pH})} \tag{6}$$

S^{unprot} values computed across the entire pH range, were then fitted to a generalized version of the Henderson-Hasselbalch (HH) formula[98] to obtain a single pK_a value:

$$S^{\text{unprot}}(\text{pH}) = \frac{1}{1+10^{-n(pH-pKa)}} \tag{7}$$

To illustrate the effect of coupled titrating residues, CPHMD$^{\text{MS}\lambda\text{D}}$ simulations on several dipeptides (see Supporting Information for discussion) were also performed. For these dipeptide simulations, the protonation state statistics for a specific residue may not be associated with the titrating residue, because the symmetry of the system may render the environment around each titrating residue to be similar. Therefore, the pK_a calculation has to be performed using a modified version of Eq. (7), where the combined S^{unprot} ratio for all *i* residues is fitted to the following equation:

$$\sum_i^N S_i^{\text{unprot}}(\text{pH}) = \sum_i^N \frac{1}{1+10^{-(pH-pKa_i)}} \tag{8}$$

In this study, the reported pK_a value and its uncertainty correspond to the mean and standard deviation calculated from three sets of independent runs.

RESULTS & DISCUSSION

Optimization of model potential parameters for two-state titrations

As with the previous implementation of CPHMD$^{\text{MS}\lambda\text{D}}$ for nucleic acids,[83] we used the free energy of deprotonation as the fixed biasing potential (F^{fixed}) in our simulation. The free energy of deprotonation was calculated for each isolated model compound embedded in explicit solvent using traditional λ-dynamics at zero ionic strength. In order to facilitate transitions between the two protonation states, we optimized the force constant (k_{bias}) on the variable biasing potential (F^{var}) that was applied to each model compound, and targeted to achieve a maximal value of the transition rate in λ-space (i.e., titration coordinate sampling), while maintaining a high fraction of physical ligands. The optimized parameters for the model potentials are reported in Table I. Calculation of the sampling statistics (see Supporting Information Table S1) indicates that the fraction of physical states was maintained at ∼70% and transitions in λ-space were ∼50 transitions/ns. The sampling properties of our model amino acids are comparable to previous work performed on model nucleosides in explicit solvent.[83]

Next, we performed a two-state titration simulation, where only two titrating states (protonated and unprotonated) were simulated, and tautomers for each protonation state were not explicitly modeled. The titration curves for our model compounds are illustrated in Figure

Table I
Parameters for the Model Potential for Two-State Titrations

		F^{var} (kcal/mol)	
Residue	$\Delta G_{protonation}$ (kcal/mol)	k_{bias}	Ref pK_a
Asp	43.71	34.00	4.00
Glu	46.00	34.25	4.40
His-δ	−3.58	26.00	7.00
His-ε	−12.26	26.00	6.60
Lys	−23.02	29.50	10.40

1. The calculated pK_a of aspartic and glutamic acid for a two-state titration (i.e., without proton tautomerism) was 4.1 and 4.3 pK_a respectively, which is within ±0.1 pK_a units from their experimental pK_a values of 4.0 and 4.4 respectively.[99] For the two-state titration of histidine, where either Nδ or Nε was titrated, the pK_a values obtained were 6.7 and 7.0 respectively,[100] which is identical to their experimental pK_a values. Finally, the calculated pK_a of lysine was 10.2, which is in close agreement with the experimental pK_a value of 10.4.[99] The excellent agreement between our model compounds calculated pK_a values and their experimental values indicate that the sampling of titration coordinates in our CPHMD$^{MS\lambda D}$ simulations was sufficient to yield accurate results.

Optimization of model potential parameters for three-state titrations

The original form of the F^{var} potential assumed the existence of only two states. When accounting for proton tautomerism and thus three states, the original form was not suitable because it frequently sampled an intermediate state of the two tautomers. This intermediate state is typically characterized by $\lambda_{\alpha,1} \approx 0$, $\lambda_{\alpha,2} \approx 0.5$ and $\lambda_{\alpha,3} \approx 0.5$, which corresponds to a half proton on both the Nδ and Nϵ protonation sites (using His as an example), and this represents an unphysical state whose sampling should be minimized. The existence of the intermediate state can be rationalized by considering that the free energy barrier for conversion between the two protonation states would be larger than the conversion between the two tautomers, as in the former process there is a change in the net charge of the system and a greater reorganization of the distribution of partial charges. The combined functional form of the original F^{var} potential that uses the same 0.8 cutoff in the definition of physical protonation states (see Methods section) as expressed in Eq. (9), where $\lambda_{\alpha,1}$, $\lambda_{\alpha,2}$ and $\lambda_{\alpha,3}$ denote the alchemical scaling factors associated with each of the three states for some residue α, does not account for the uneven barrier height of the different alchemical reactions.

$$F_{\alpha}^{var} = k_1\left(\lambda_{\alpha,1}-0.8\right)^2 + k_1\left(\lambda_{\alpha,2}-0.8\right)^2 + k_1\left(\lambda_{\alpha,3}-0.8\right)^2 \quad (9)$$

To avoid the intermediate tautomeric states, we modified the existing F^{var} potential by including additional cross terms (k_2 expressions) to account for uneven barrier heights, and a final term (k_3 expressions) was added to ensure that the relative free energy of the end-states were not altered. The resulting functional form as outlined in Eq. (10) results in a more versatile biasing potential that is suited to address the asymmetry of the potential energy surface associated with changes in both protonation and tautomeric states.

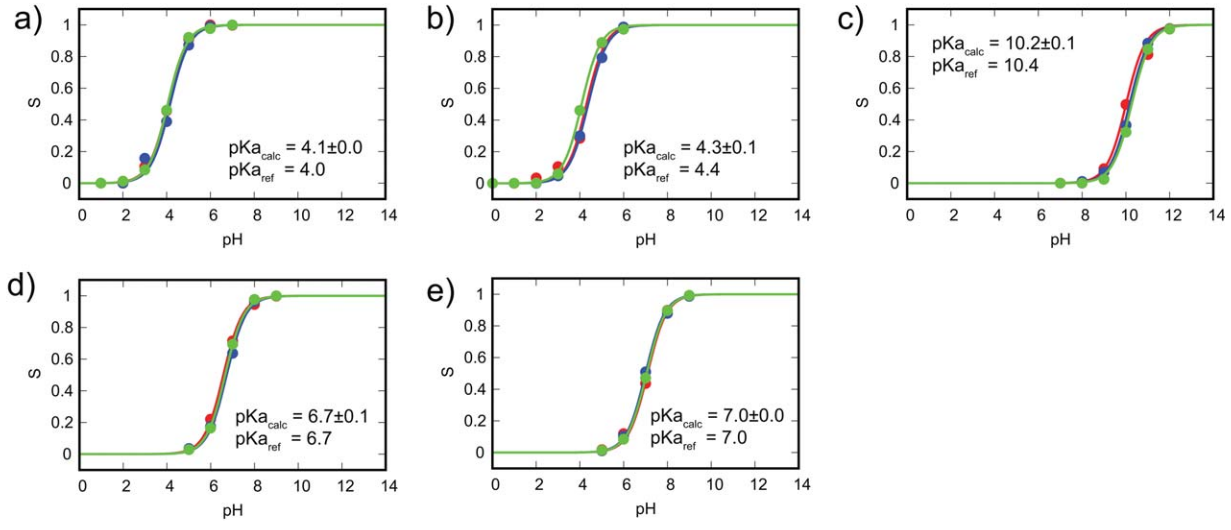

Figure 1

Titration curve of model compounds: (**a**) aspartic acid, (**b**) glutamic acid, (**c**) lysine, (**d**) histidine-δ, and (*e*) histidine-ϵ. Calculated pK_a values of model compounds are in excellent agreement with experimental pK_a values. Colors represent the results from the triplicate runs.

Table II

Parameters for the Model Potential for Three-State Titrations

		F^{var} (kcal/mol)			
Residue	$\Delta G_{protonation}$ (kcal/mol)	k_1	k_2	k_3	Ref. pK_a
Asp-T	43.30	−16.5	18.5	−18.5	4.00
Glu-T	45.59	−16.0	18.5	−18.5	4.40
His-T	−3.58/−12.26	8.0	6.0	−6.0	6.45

Table III

Sampling Characteristics of Three-State Titration Simulations Performed at the pH Closest to the Model Compound's pK_a Value

		Normal MD		pH-REX	
Residue	pH	FPS	Transition (ns^{-1})	FPS	Transition (ns^{-1})
Asp-T	4	0.78 ± 0.01	50 ± 1	0.78 ± 0.01	294 ± 16
Glu-T	4	0.76 ± 0.01	46 ± 7	0.75 ± 0.00	322 ± 21
His-T	7	0.81 ± 0.02	60 ± 6	0.81 ± 0.01	298 ± 12

Using pH-REX greatly accelerates sampling of titration coordinates with minimal loss in the fraction of physical states (FPS).

$$F_{\alpha}^{var} = k_1\left(\lambda_{\alpha,1}-0.8\right)^2 + k_1\left(\lambda_{\alpha,2}-0.8\right)^2 + k_1\left(\lambda_{\alpha,3}-0.8\right)^2 + k_2\left(\lambda_{\alpha,1}-\lambda_{\alpha,2}\right)^2 k_2\left(\lambda_{\alpha,1}-\lambda_{\alpha,3}\right)^2 - k_3\left(\lambda_{\alpha,2}\right) - k_3\left(\lambda_{\alpha,3}\right) \tag{10}$$

An iterative grid search strategy was used in testing various combinations of the force constants (k_1, k_2, k_3), and the optimal combination is reported in Table II. As illustrated in Figure 2, which shows the time-evolution of λ, all three end states for the model compounds were well sampled. Calculation of the sampling statistics as summarized in Table III indicates that the fraction of physical states was maintained above 70%, confirming that the modified F^{var} potential does not trap λ in an unphysical intermediate state. The transitions in λ-space were ~50 transitions/ns, which is comparable to the statistics obtained from two-state titrations of the model compounds.

While the sampling efficiency in λ-space of model compounds allows us to reproduce the pK_a values of the model compounds, the transition rate is nevertheless limited to ~50 transitions/ns. In our previous evaluation of explicit solvent CPHMD$^{MS\lambda D}$ simulations of larger nucleic acid structures, slower pK_a convergence was observed,[84] and it is likely that protein systems will encounter similar issues as well. The sampling of titration and spatial coordinates can be accelerated using a pH-REX sampling strategy.[85] Therefore, we have applied pH-REX sampling, and as illustrated in Table III, it resulted in a sixfold improvement in λ-space sampling of model compounds with effectively no loss in the fraction of physical states. As pH-REX sampling confers significant improvement over straightforward MD simulations and requires negligible overhead in terms of computational cost, the results presented in the subsequent sections are obtained from pH-REX CPHMD$^{MS\lambda D}$ simulations unless specified otherwise.

We performed a three-state titration on the model compounds, where alchemical transformations across different protonation states and across different tautomers of the same protonation state were explicitly modeled. The tautomeric titrations of aspartic and glutamic yielded a pK_a of 4.4 and 4.8 respectively, which matches well with the macroscopic pK_a of 4.35 and 4.70 when the double degeneracy of the protonated states is taken into account.[62] However, since the experimental pK_a measured does not distinguish between the tautomeric forms, we recalibrated the fixed biasing potential in our CPHMD$^{MS\lambda D}$ simulations to reproduce the experimentally measured macroscopic pK_a values. This was

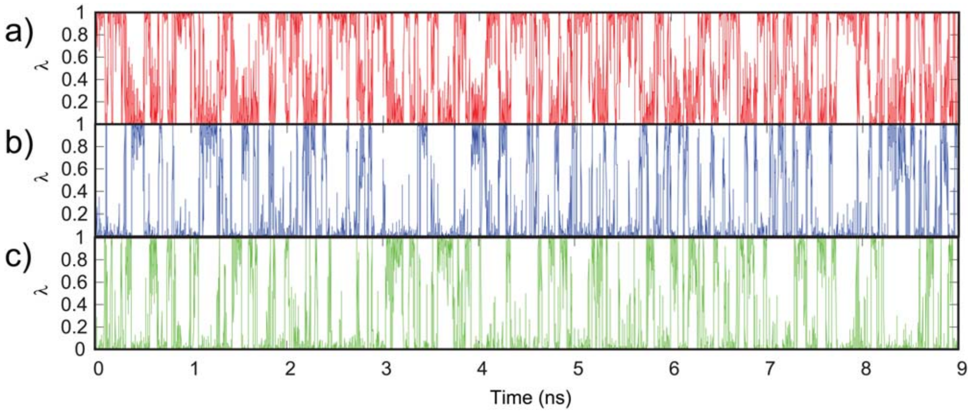

Figure 2

Titration coordinate transitions of aspartic acid at pH 4 for (**a**) unprotonated state, (**b**) protonated tautomer #1, and (**c**) protonated tautomer #2 shows that the physical end states are well sampled.

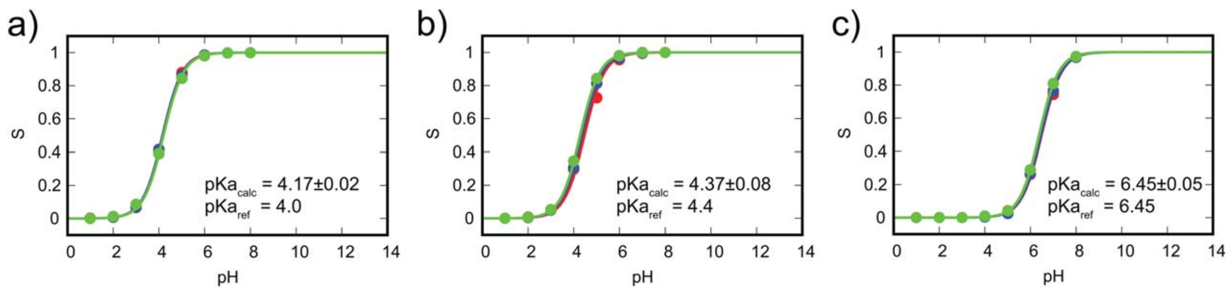

Figure 3

Titration curve of model compounds with proton tautomerism: (**a**) aspartic acid, (**b**) glutamic acid, and (**c**) histidine. Calculated pK_a values of model compounds show excellent agreement with experimental pK_a values. Colors represent the results from the triplicate runs.

achieved by reducing the biasing potential at pH = pK_a by $k_b T\ln(2) = 0.41$ kcal/mol, which accounts for the degeneracy of the tautomeric protonated states. Our approach is different from that of Khandogin and Brooks,[62] where a post-correction factor of 0.3 pK_a units was applied to tautomeric residues. However, the net result in both approaches is the same, in the sense that the final pK_a value calculated accounts for tautomer degeneracy. The titration curves for our model compounds with proton tautomerism are illustrated in Figure 3. The calculated pK_a of aspartic and glutamic acid was 4.17 and 4.37 respectively, which is good agreement with experimental pK_a values. For histidine tautomeric titration, no re-calibration was performed because the pK_a measured by experiments were microscopic pK_a associated with the titration at the Nϵ and Nδ sites, and the fixed biasing potential applied to each tautomer was identical to those used in the 2-state titration setup. Our calculated pK_a for the histidine tautomer was 6.45, which is identical to the expected macroscopic pK_a value of 6.45.[62]

pK_a calculations of proteins: Hen-egg white lysozyme (HEWL)

The HEWL protein is a well-studied protein system that contains the three most common titrating residues (Asp, Glu, His) with microscopic pK_a values for each residue that have been measured in a number of experimental studies.[101–106] It is perhaps the closest thing to a "universal benchmark" system that has been evaluated by numerous CPHMD implementations over the years.[43,48,49,51,63,75,107] To the best of our knowledge, all existing "pure" explicit solvent CPHMD simulations reported in the literature have only been demonstrated on small peptide compounds[80] and simple organic molecules.[108] We performed a 20 ns pH-REX CPHMD$^{MS\lambda D}$ simulation of HEWL, which is the first example of explicit solvent CPHMD simulation on a full protein to be reported.

pK_a calculations over 5 ns interval segments of our pH-REX CPHMD$^{MS\lambda D}$ trajectory show that good convergence is achieved within 20 ns (Supporting Information Table S3). The difference in pK_a values across our triplicate runs is small, typically between 0.2 and 0.3 pK_a units, demonstrating that our results are robust and reproducible. The accuracy of our calculated pK_a values are then compared to experimental measurements from consensus NMR titrations.[106] As summarized in Table IV and Supporting Information Figure S3, the calculated pK_a values are in good agreement with experiment, with a root-mean-square-error (RMSE) of 0.85 pK_a units and an average unsigned error (AUE) of 0.68 pK_a units. Webb *et al.* previously estimated that experimental pK_a values reported in the literature on average may vary by 0.5 pK_a units depending on the experimental method and/or protocol used to make the measurements.[106] This suggests that the accuracy of our pH-REX CPHMD$^{MS\lambda D}$ simulations are approaching the uncertainty of experimental pK_a values. Next, we identified the residues that had errors in their calculated pK_a values, which we defined as having more than 1.0 pK_a unit difference between calculated and experimental values. Asp-119 was underpredicted by −1.9 pK_a units, which suggests that the unprotonated state is overstabilized in our simulations. Analysis of its microenvironment indicates that persistent hydrogen bond interactions between the carboxylic oxygens of Asp-119 and the amide backbone hydrogen of Gln-121 and Ala-122 were present even in a low pH environment (Supporting Information Fig. S4), which accounts for the extra stabilization of the unprotonated state of Asp-119. Similar underprediction of Asp pK_a values has been documented in other CPHMD work, where salt bridge interactions were responsible.[49] When non-salt-bridge configurations were sampled, it resulted in more accurate pK_a results.[49] This suggests that the apparent error in the Asp-119 pK_a value could be a sampling issue, and more extensive sampling or more aggressive sampling methods may be required when dealing with residues that are "locked" to their initial conformation by strong interactions like hydrogen bonds or salt bridges.

We compared the performance of explicit solvent pH-REX CPHMD$^{MS\lambda D}$ simulations to CPHMD models implemented in other solvation models. A number of

Table IV

pK_a Values of HEWL Calculated Using Implicit and Hybrid Solvent pH-REX CPHMD Simulations as Reported by Wallace and Shen,[75] Compared with pK_a Values Calculated Using Explicit Solvent pH-REX CPHMD$^{MS\lambda D}$ Simulations in this Work

		Implicit CPHMD		Hybrid CPHMD		Explicit CPHMD$^{MS\lambda D}$	
Residue	Exp pK_a	pK_a	Error	pK_a	Error	pK_a	Error
GLU-7	2.6 ± 0.2	2.6 ± 0.1	0.0	2.7 ± 0.0	0.1	2.7 ± 0.1	0.1
HIS-15	5.5 ± 0.2	5.3 ± 0.5	−0.2	6.6 ± 0.1	1.1	6.0 ± 0.2	0.5
ASP-18	2.8 ± 0.3	2.9 ± 0.0	0.1	3.1 ± 0.1	0.3	2.1 ± 0.2	−0.7
GLU-35	6.1 ± 0.4	4.4 ± 0.2	−1.8	7.2 ± 0.2	1.1	7.0 ± 0.3	0.9
ASP-48	1.4 ± 0.2	2.8 ± 0.2	1.4	1.6 ± 0.5	0.2	1.3 ± 0.0	−0.1
ASP-52	3.6 ± 0.3	4.6 ± 0.0	1.0	2.9 ± 0.1	−0.7	4.5 ± 0.3	0.9
ASP-66	1.2 ± 0.2	1.2 ± 0.4	−0.1	1.5 ± 0.6	0.3	1.5 ± 0.1	0.3
ASP-87	2.2 ± 0.1	2.0 ± 0.1	−0.2	1.5 ± 0.4	−0.7	1.3 ± 0.0	−0.9
ASP-101	4.5 ± 0.1	3.3 ± 0.3	−1.2	3.0 ± 0.1	−1.5	5.1 ± 0.5	0.6
ASP-119	3.5 ± 0.3	2.5 ± 0.1	−1.1	2.9 ± 0.1	−0.7	1.6 ± 0.0	−1.9
RMSE		0.94		0.80		0.84	
AUE		0.70		0.66		0.68	

Calculated pK_a values with error greater than 1.0 pK_a unit relative to experimental values based on consensus NMR titrations[106] are identified in red.

CPHMD variations have been implemented in AMBER[48] and GROMACS.[43] However, they will not be included our analysis as deconvoluting the effects originating from force field differences to those arising from solvation model differences is not straightforward. Instead, we will focus our analysis on CPHMD variations implemented in CHARMM. The original CPHMD in CHARMM was implemented with a GB implicit solvent model,[61] and we have used the HEWL pK_a values reported by Wallace and Shen for comparison.[75] Since that work was reported using a pH-REX sampling strategy, we have also eliminated the effects of using different sampling strategies. At the time of writing, there is no "pure" explicit solvent CPHMD based on the CHARMM force field that has been tested on the HEWL protein. However, a close comparison can be made with Shen's hybrid solvent CPHMD model.[75] The key methodological difference between explicit and hybrid solvent models is that the evaluation of free energies of deprotonation and the forces on the fictitious λ particles that govern the titration coordinates are calculated using a GB implicit solvent model in Shen's hybrid solvent CPHMD model, whereas in our explicit solvent CPHMD$^{MS\lambda D}$ model there is no use of the GB implicit solvent model in any part of the calculation. Unfortunately, the use of such hybrid sampling means there is no clear Hamiltonian for this system and correspondence to results from any specific statistical mechanical approach cannot be demonstrated. Lastly, the sampling of titration coordinates in implicit solvent is typically ~2000 transitions/ns,[51] which is an order of magnitude higher than those obtained in our explicit solvent simulations. Therefore, to compensate for the differential sampling speed associated with different solvent models, we compared the results of our 20 ns pH-REX CPHMD$^{MS\lambda D}$ trajectories to the previously reported 2 ns pH-REX trajectories that uses the implicit and hybrid CPHMD model.

As summarized in Table IV, in terms of overall pK_a predictive performance, our explicit solvent CPHMD$^{MS\lambda D}$ results had a RMSE error of 0.84 pK_a units. This is an improvement from the results obtained using implicit solvent CPHMD (RMSE = 0.94), and our model performance is close to that of the hybrid solvent CPHMD (RMSE = 0.80). A similar trend was also noted using alternative error metrics, such as the average unsigned error (AUE). We then identified the number of residues that had errors of more than 1.0 pK_a unit relative to experimental values. Our explicit solvent CPHMD$^{MS\lambda D}$ model had only 1 such residue (i.e., Asp-119) compared to the implicit and hybrid solvent CPHMD models which had five and three residues, respectively. Notable improvements in moving from a hybrid solvent to a "pure" explicit solvent model can be observed in His-15, where the overestimation of its pK_a value is reduced from 1.1 to 0.5 pK_a units. Similarly, the hybrid solvent CPHMD model incorrectly predicted the direction of pK_a shift for residue Asp-101, whereas the explicit solvent CPHMD$^{MS\lambda D}$ model not only predicted the right direction of pK_a shift, but the magnitude of error was also smaller (−1.5 vs. +0.6). Our findings suggest that when corrected for differences in titrating coordinates sampling, the explicit solvent CPHMD$^{MS\lambda D}$ model produces more accurate pK_a predictions than the original implicit solvent CPHMD.

pK_a calculations of proteins: Other proteins, BBL, and NTL9

Lastly, to demonstrate that the pK_a calculations obtained from the CPHMD$^{MS\lambda D}$ framework for proteins is not specific to HEWL protein, we performed pK_a calculations on two additional proteins, the BBL and NTL9 protein. Given that we have only investigated a single His residue in a protein environment, for BBL we only

Table V
pK_a Values of BBL and NTL9 Calculated Using Explicit Solvent pH-REX CPHMD$^{MS\lambda D}$ Simulations in this Work

		Explicit CPHMD$^{MS\lambda D}$	
Residue	Exp pK_a	pK_a	Error
BBL			
HIS-142	6.5	6.6 ± 0.1	0.1
HIS-166	5.4	4.8 ± 0.0	−0.6
NTL9			
ASP-8	3.0	1.5 ± 0.1	−1.5
GLU-17	3.6	4.0 ± 0.5	0.4
ASP-23	3.1	3.7 ± 0.2	0.6
GLU-38	4.0	3.9 ± 0.2	−0.1
GLU-48	4.2	3.4 ± 0.3	−0.8
GLU-54	4.2	3.6 ± 0.2	−0.6
RMSE			0.72
AUE			0.59

Calculated pK_a values with error greater than 1.0 pK_a unit relative to experimental values[109,110] based are identified in red.

titrated the two His residues. NTL9 has no His residues, and the Glu and Asp residues that have experimental pK_a measurements were titrated. As summarized in Table V and Supporting Information Figure S5 the calculated pK_a values are reasonably accurate (RMSE = 0.72, AUE = 0.59).[109,110] From the experimental data, most of the residues titrate close to the pK_a of their reference compounds, but two residues had more than a 1.0 pH unit shift. For His-166 of BBL, the residue is buried and its experimental pK_a is 5.4. For Asp-8 of NTL9, its experimental pK_a of 3.0 can be traced to the salt bridge interactions it forms with the amide backbone of adjacent residues. Our calculated pK_a values demonstrate a similar downward shift, although in both cases the extent of the shift tends to be overestimated. We suggest that this overestimation may be due to the lack of sampling stemming from the shorter 3 to 5 ns simulations performed for these systems. In other proteins like staphylococcal nuclease, residues with shifted pK_a values are known to undergo local conformational changes,[26,28] and sampling these states will be required to improve the accuracy of pK_a calculations. Together with our observations for Asp-119 in HEWL, our work suggests that while short pH-REX CPHMD$^{MS\lambda D}$ simulations are capable of reproducing experimental pK_a values of most protein residues, accurate reproduction of highly shifted pK_a values (e.g., buried charged residues) or those involving salt-bridge or similarly strong interactions remains a challenge that may be better addressed with more aggressive conformational sampling techniques.

CONCLUSION

In conclusion, we have demonstrated the use of the constant pH molecular dynamics framework based on multi-site λ-dynamics (CPHMD$^{MS\lambda D}$) to simulate realistic pH-dependent properties of proteins. In the CPHMD$^{MS\lambda D}$ framework, we performed seamless alchemical transitions between protonation and tautomeric states using multi-site λ-dynamics, and designed a novel biasing potential to ensure that only the physical end-states are predominantly sampled. Then, we applied explicit solvent CPHMD$^{MS\lambda D}$ simulations to the proteins HEWL, BBL and NTL9, which are the first examples of a "pure" explicit solvent CPHMD on full protein systems to be reported. Our pK_a calculations for HEWL protein are in excellent agreement with experimental values, with a RMSE of 0.84 pK_a units, and this is close to the uncertainty of 0.50 pK_a units associated with experimental measurements. Our pK_a calculations on the other model protein systems, BBL and NTL9 also provide similarly good agreement with experiments. In addition, comparison with implicit solvent CPHMD shows that explicit solvent CPHMD$^{MS\lambda D}$ produces results that are more accurate, reducing the number of residues with large errors in their pK_a predictions from 5 to 1. With the development of explicit solvent CPHMD$^{MS\lambda D}$ for proteins, it will finally allow us to confidently address questions related to pH-dependent properties of membrane proteins and ion channels, where discrete representation of ions and water is important. Coupled with the explicit solvent CPHMD$^{MS\lambda D}$ framework for nucleic acids, accurate modeling of pH-dependent properties for all major classes of biomolecules—proteins, nucleic acids, and even protein-nucleic acid complexes is now a reality.

ACKNOWLEDGMENTS

The authors thank Dr. Afra Panahi for critically reviewing the manuscript and providing helpful comments.

REFERENCES

1. Warshel A. Calculations of enzymatic-reactions—calculations of pKa, proton-transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 1981;20:3167–3177.
2. Harris TK, Turner GJ. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. IUBMB Life 2002;53:85–98.
3. Nielsen JE, Mccammon JA. Calculating pKa values in enzyme active sites. Protein Sci 2003;12:1894–1901.
4. Demchuk E, Genick UK, Woo TT, Getzoff ED, Bashford D. Protonation states and pH titration in the photocycle of photoactive yellow protein. Biochemistry 2000;39:1100–1113.
5. Dillet V, Dyson HJ, Bashford D. Calculations of electrostatic interactions and pKas in the active site of Escherichia coli thioredoxin. Biochemistry 1998;37:10298–10306.
6. Wilcox JL, Ahluwalia AK, Bevilacqua PC. Charged nucleobases and their potential for RNA catalysis. Acc Chem Res 2011;44:1270–1279.
7. Krishnamurthy R. Role of pK(a) of nucleobases in the origins of chemical evolution. Acc Chem Res 2012;45:2035–2044.
8. Shih IH, Been MD. Involvement of a cytosine side chain in proton transfer in the rate-determining step of ribozyme self-cleavage. Proc Natl Acad Sci USA 2001;98:1489–1494.

9. Ke A, Zhou K, Ding F, Cate JH, Doudna JA. A conformational switch controls hepatitis delta virus ribozyme catalysis. Nature 2004;429:201–205.
10. Ravindranathan S, Butcher SE, Feigon J. Adenine protonation in domain B of the hairpin ribozyme. Biochemistry 2000;39:16026–16032.
11. Ryder SP, Oyelere AK, Padilla JL, Klostermeier D, Millar DP, Strobel SA. Investigation of adenosine base ionization in the hairpin ribozyme by nucleotide analog interference mapping. RNA 2001;7:1454–1463.
12. Bierzynski A, Kim PS, Baldwin RL. A salt bridge stabilizes the helix formed by isolated C-peptide of Rnase-A. Proc Natl Acad Sci USA 1982;79:2470–2474.
13. Shoemaker KR, Kim PS, Brems DN, Marqusee S, York EJ, Chaiken IM, Stewart JM, Baldwin RL. Nature of the charged-group effect on the stability of the C-peptide helix. Proc Natl Acad Sci USA 1985;82:2349–2353.
14. Schaefer M, Van Vlijmen HWT, Karplus M. Electrostatic contributions to molecular free energies in solution. Adv Protein Chem 1998;51:1–57.
15. Kelly JW. Alternative conformations of amyloidogenic proteins govern their behavior. Curr Opin Struct Biol 1996;6:11–17.
16. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol 2000;10:153–159.
17. Warshel A. Electrostatic basis of structure-function correlation in proteins. Acc Chem Res 1981;14:284–290.
18. Hunenberger PH, Helms V, Narayana N, Taylor SS, McCammon JA. Determinants of ligand binding to cAMP-dependent protein kinase. Biochemistry 1999;38:2358–2366.
19. Houck-Loomis B, Durney MA, Salguero C, Shankar N, Nagle JM, Goff SP, D'Souza VM. An equilibrium-dependent retroviral mRNA switch regulates translational recoding. Nature 2011;480: 561–U193.
20. Webb BA, Chimenti M, Jacobson MP, Barber DL. Dysregulated pH: a perfect storm for cancer progression. Nat Rev Cancer 2011; 11:671–677.
21. Howell EE, Villafranca JE, Warren MS, Oatley SJ, Kraut J. Functional-role of aspartic acid-27 in dihydrofolate-reductase revealed by mutagenesis. Science 1986;231:1123–1128.
22. Rastogi VK, Girvin ME. Structural changes linked to proton translocation by subunit c of the ATP synthase. Nature 1999;402:263–268.
23. Reiter NJ, Blad H, Abildgaard F, Butcher SE. Dynamics in the U6 RNA intramolecular stem-loop: a base flipping conformational change. Biochemistry 2004;43:13739–13747.
24. Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza hemagglutinin at the pH of membrane-fusion. Nature 1994;371:37–43.
25. Harms MJ, Schlessman JL, Sue GR, Garcia-Moreno B. Arginine residues at internal positions in a protein are always charged. Proc Natl Acad Sci USA 2011;108:18954–18959.
26. Isom DG, Castaneda CA, Cannon BR, Garcia-Moreno B. Large shifts in pKa values of lysine residues buried inside a protein. Proc Natl Acad Sci USA 2011;108:5260–5265.
27. Pey AL, Rodriguez-Larrea D, Gavira JA, Garcia-Moreno B, Sanchez-Ruiz JM. Modulation of buried ionizable groups in proteins with engineered surface charge. J Am Chem Soc 2010;132: 1218–1219.
28. Isom DG, Cannon BR, Castaneda CA, Robinson A, Garcia-Moreno B. High tolerance for ionizable residues in the hydrophobic interior of proteins. Proc Natl Acad Sci USA 2008;105:17784–17788.
29. Nielsen JE, Gunner MR, Garcia-Moreno BE. The pKa Cooperative: a collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. Proteins 2011;79: 3249–3259.
30. Alexov E, Mehler EL, Baker N, Baptista AM, Huang Y, Milletti F, Nielsen JE, Farrell D, Carstensen T, Olsson MH, Shen JK, Warwicker J, Williams S, Word JM. Progress in the prediction of pKa values in proteins. Proteins 2011;79:3260–3275.
31. Bashford D. Macroscopic electrostatic models for protonation states in proteins. Front Biosci 2004;9:1082–1099.
32. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
33. You TJ, Bashford D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. Biophys J 1995;69:1721–1733.
34. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. Biophys J 2002;83:1731–1748.
35. Russell ST, Warshel A. Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. J Mol Biol 1985;185:389–404.
36. Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the polaris and enzymix programs. J Comput Chem 1993;14:161–185.
37. Warshel A, Sussman F, King G. Free-energy of charges in solvated proteins—microscopic calculations using a reversible charging process. Biochemistry 1986;25:8368–8372.
38. Sham YY, Chu ZT, Warshel A. Consistent calculations of pKa's of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B 1997;101:4458–4472.
39. Mongan J, Case DA. Biomolecular simulations at constant pH. Curr Opin Struct Biol 2005;15:157–163.
40. Burgi R, Kollman PA, van Gunsteren WF. Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. Proteins 2002;47:469–480.
41. Baptista AM, Teixeira VH, Soares CM. Constant-pH molecular dynamics using stochastic titration. J Chem Phys 2002;117:4184–4200.
42. Machuqueiro M, Baptista AM. Constant-pH molecular dynamics with ionic strength effects: protonation-conformation coupling in decalysine. J Phys Chem B 2006;110:2927–2933.
43. Baptista AM, Machuqueiro M. Acidic range titration of HEWL using a constant-pH molecular dynamics method. Proteins 2008; 72:289–298.
44. Im WP, Lee MS, Brooks CL, III. Generalized born model with a simple smoothing function. J Comput Chem 2003;24:1691–1702.
45. Chen JH, Im WP, Brooks CL, III. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. J Am Chem Soc 2006;128:3728–3736.
46. .Dlugosz M, Antosiewicz JM. Constant-pH molecular dynamics simulations: a test case of succinic acid. Chem Phys 2004;302:161–170.
47. Dlugosz M, Antosiewicz JM, Robertson AD. Constant-pH molecular dynamics study of protonation-structure relationship in a heptapeptide derived from ovomucoid third domain. Phys Rev E 2004;69:021915.
48. Mongan J, Case DA, McCammon JA. Constant pH molecular dynamics in generalized born implicit solvent. J Comput Chem 2004;25:2038–2048.
49. Williams SL, de Oliveira CA, McCammon JA. Coupling constant pH molecular dynamics with accelerated molecular dynamics. J Chem Theory Comput 2010;6:560–568.
50. Meng Y, Roitberg AE. Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. J Chem Theory Comput 2010;6:1401–1412.
51. Swails JM, Roitberg AE. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. J Chem Theory Comput 2012;8: 4393–4404.

52. Sabri Dashti D, Meng Y, Roitberg AE. pH-replica exchange molecular dynamics in proteins using a discrete protonation method. J Phys Chem B 2012;116:8805–8811.
53. Messer BM, Roca M, Chu ZT, Vicatos S, Kilshtain AV, Warshel A. Multiscale simulations of protein landscapes: using coarse-grained models as reference potentials to full explicit models. Proteins 2010;78:1212–1227.
54. Olsson MHM, Warshel A. Monte Carlo simulations of proton pumps: on the working principles of the biological valve that controls proton pumping in cytochrome c oxidase. Proc Natl Acad Sci USA 2006;103:6500–6505.
55. Aaqvist J, Warshel A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. Chem Rev 1993;93:2523–2544.
56. Baptista AM, Martel PJ, Petersen SB. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. Proteins 1997;27:523.
57. Borjesson U, Hunenberger PH. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. J Chem Phys 2001;114:9706–9719.
58. Kong X, Brooks CL, III. Lambda-dynamics-a new approach to free-energy calculations. J Chem Phys 1996;105:2414–2423.
59. Knight JL, Brooks CL, III. Lambda-dynamics free energy simulation methods. J Comput Chem 2009;30:1692–1700.
60. Guo Z, Brooks CL, III, Kong X. Efficient and flexible algorithm for free energy calculations using the λ-dynamics approach. J Phys Chem B 1998;102:2032–2036.
61. Lee MS, Salsbury FR, Brooks CL, III. Constant-pH molecular dynamics using continuous titration coordinates. Proteins 2004;56: 738–752.
62. Khandogin J, Brooks CL, III. Constant pH molecular dynamics with proton tautomerism. Biophys J 2005;89:141–157.
63. Khandogin J, Brooks CL, III. Toward the accurate first-principles prediction of ionization equilibria in proteins. Biochemistry 2006; 45:9363–9373.
64. Khandogin J, Chen J, Brooks CL, III. Exploring atomistic details of pH-dependent peptide folding. Proc Natl Acad Sci USA 2006; 103:18546–18550.
65. Khandogin J, Raleigh DP, Brooks CL, III. Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen-bonded network. J Am Chem Soc 2007;129: 3056–3057.
66. Khandogin J, Brooks CL, III. Linking folding with aggregation in Alzheimer's beta-amyloid peptides. Proc Natl Acad Sci USA 2007; 104:16880–16885.
67. Zhang BW, Brunetti L, Brooks CL, III. Probing pH-dependent dissociation of HdeA dimers. J Am Chem Soc 2011;133:19393–19398.
68. Shen JK. Uncovering specific electrostatic interactions in the denatured states of proteins. Biophys J 2010;99:924–932.
69. Wallace JA, Shen JK. Unraveling a trap-and-trigger mechanism in the pH-sensitive self-assembly of spider silk proteins. J Phys Chem Lett 2012;3:658–662.
70. Law SM, Zhang BW, Brooks CL, III. pH-sensitive residues in the p19 RNA silencing suppressor protein from carnation Italian ringspot virus affect siRNA binding stability. Protein Sci 2013;22:595–604.
71. Dlugosz M, Antosiewicz JM. Effects of solute-solvent proton exchange on polypeptide chain dynamics: a constant-pH molecular dynamics study. J Phys Chem B 2005;109:13777–13784.
72. Machuqueiro M, Baptista AM. The pH-dependent conformational states of kyotorphin: a constant-pH molecular dynamics study. Biophys J 2007;92:1836–1845.
73. Campos SR, Machuqueiro M, Baptista AM. Constant-pH molecular dynamics simulations reveal a beta-rich form of the human prion protein. J Phys Chem B 2010;114:12692–12700.
74. Arthur EJ, Yesselman JD, Brooks CL, III. Predicting extreme pK(a) shifts in staphylococcal nuclease mutants with constant pH molecular dynamics. Proteins 2011;79:3276–3286.
75. Wallace JA, Shen JK. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. J Chem Theory Comput 2011;7:2617–2629.
76. Wang WZ, Chu XP, Li MH, Seeds J, Simon RP, Xiong ZG. Modulation of acid-sensing ion channel currents, acid-induced increase of intracellular Ca2+, and acidosis-mediated neuronal injury by intracellular pH. J Biol Chem 2006;281:29369–29378.
77. Hesselager M, Timmermann DB, Ahring PK. pH dependency and desensitization kinetics of heterologously expressed combinations of acid-sensing ion channel subunits. J Biol Chem 2004;279: 11006–11015.
78. Berdiev BK, Mapstone TB, Markert JM, Gillespie GY, Lockhart J, Fuller CM, Benos DJ. pH alterations "reset" Ca2+ sensitivity of brain Na+ channel 2, a degenerin/epithelial Na+ ion channel, in planar lipid bilayers. J Biol Chem 2001;276:38755–38761.
79. Damaghi M, Bippes C, Koster S, Yildiz O, Mari SA, Kuhlbrandt W, Muller DJ. pH-dependent interactions guide the folding and gate the transmembrane pore of the beta-barrel membrane protein OmpG. J Mol Biol 2010;397:878–882.
80. Donnini S, Tegeler F, Groenhof G, Grubmuller H. Constant pH molecular dynamics in explicit solvent with lambda-dynamics. J Chem Theory Comput 2011;7:1962–1978.
81. Knight JL, Brooks CL, III. Applying efficient implicit nongeometric constraints in alchemical free energy simulations. J Comput Chem 2011;32:3423–3432.
82. Knight JL, Brooks CL, III. Multisite λ dynamics for simulated structure–activity relationship studies. J Chem Theory Comput 2011;7:2728–2739.
83. Goh GB, Knight JL, Brooks CL, III. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. J Chem Theory Comput 2012;8:36–46.
84. Goh GB, Knight JL, Brooks CL, III. pH-dependent dynamics of complex RNA macromolecules. J Chem Theory Comput 2013;9: 935–943.
85. Goh GB, Knight JL, Brooks CL, III. Toward accurate prediction of the protonation equilibrium of nucleic acids. J Phys Chem Lett 2013;4:760–766.
86. Zheng L, Chen M, Yang W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. Proc Natl Acad Sci USA 2008;105:20227–20232.
87. Shi CY, Wallace JA, Shen JK. Thermodynamic coupling of protonation and conformational equilibria in proteins: theory and simulation. Biophys J 2012;102:1590–1597.
88. Brooks BR, Brooks CL, III, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30: 1545–1614.
89. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
90. Feig M, Karanicolas J, Brooks CL, III. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graphics Modell 2004;22:377–395.
91. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular

modeling and dynamics studies of proteins. J Phys Chem B 1998; 102:3586–3616.

92. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.

93. Cheatham TE, Miller JL, Fox T, Darden TA, Kollman PA. Molecular-dynamics simulations on solvated biomolecular systems—the particle mesh ewald method leads to stable trajectories of DNA, RNA, and Proteins. J Am Chem Soc 1995;117:4193–4194.

94. Schreiber H, Steinhauser O. Cutoff size does strongly influence molecular-dynamics results on solvated polypeptides. Biochemistry 1992;31:5856–5860.

95. Steinbach PJ, Brooks BR. New spherical-cutoff methods for long-range forces in macromolecular simulation. J Comput Chem 1994; 15:667–683.

96. Beck DAC, Armen RS, Daggett V. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. Biochemistry 2005;44:609–616.

97. Norberg J, Nilsson L. On the truncation of long-range electrostatic interactions in DNA. Biophys J 2000;79:1537–1553.

98. Onufriev A, Case DA, Ullmann GM. A novel view of pH titration in biomolecules. Biochemistry 2001;40:3413–3419.

99. Nozaki Y, Tanford C. Examination of titration behavior. Methods Enzymol 1967;11:715–734.

100. Bashford D, Case DA, Dalvit C, Tennant L, Wright PE. Electrostatic calculations of side-chain pK(a) values in myoglobin and comparison with NMR data for histidines. Biochemistry 1993;32: 8045–8056.

101. Parsons SM, Raftery MA. Ionization behavior of the catalytic carboxyls of lysozyme. Effects of ionic strength. Biochemistry 1972; 11:1623–1629.

102. Kuramitsu S, Ikeda K, Hamaguchi K, Fujio H, Amano T. Ionization constants of Glu 35 and Asp 52 in hen, turkey, and human lysozymes. J Biochem 1974;76:671–683.

103. Kuramitsu S, Ikeda K, Hamaguchi K. Participation of the catalytic carboxyls, Asp 52 and Glu 35, and Asp 101 in the binding of substrate analogues to hen lysozyme. J Biochem 1975;77:291–301.

104. Takahashi T, Nakamura H, Wada A. Electrostatic forces in two lysozymes: calculations and measurements of histidine pKa values. Biopolymers 1992;32:897–909.

105. Bartik K, Redfield C, Dobson CM. Measurement of the individual Pk(a) values of acidic residues of hen and turkey lysozymes by 2-dimensional H-1-Nmr. Biophys J 1994;66:1180–1184.

106. Webb H, Tynan-Connolly BM, Lee GM, Farrell D, O'Meara F, Søndergaard CR, Teilum K, Hewage C, McIntosh LP, Nielsen JE. Remeasuring HEWL pK(a) values by NMR spectroscopy: methods, analysis, accuracy, and implications for theoretical pK(a) calculations. Proteins 2011;79:685–702.

107. Machuqueiro M, Baptista AM. Is the prediction of pKa values by constant-pH molecular dynamics being hindered by inherited problems? Proteins 2011;79:3437–3447.

108. Wallace JA, Shen JK. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. J Chem Phys 2012;137:184105.

109. Arbely E, Rutherford TJ, Sharpe TD, Ferguson N, Fersht AR. Downhill versus barrier-limited folding of BBL 1: energetic and structural perturbation effects upon protonation of a histidine of unusually low pK(a). J Mol Biol 2009;387:986–992.

110. Kuhlman B, Luisi DL, Young P, Raleigh DP. pK(a) values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. Biochemistry 1999;38: 4896–4903.